

УДК 004.94 (004.62)

Oleksandr Terentiev, Doctor of Technical Sciences, Associate Professor, Principal researcher, Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine

ORCID ID: <https://orcid.org/0000-0002-4288-1753> *e-mail*: o.terentiev@gmail.com

Volodymyr Duda, graduate student, Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine

ORCID ID: <https://orcid.org/0009-0002-4278-4635> *e-mail*: dudavolodimir@gmail.com

Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

METHOD FOR RESTORING MISSING DATA BASED ON A COMBINED EXPONENTIAL SMOOTHING MODEL

Abstract. *Data processing and analysis are often accompanied by the problem of missing values, which can significantly affect the accuracy of predictive models and decision-making. One of the main causes of missing data in energy consumption is the periodic shutdown of systems responsible for data collection and transmission. Such disruptions can lead to data loss, complicating further processing and analysis. Therefore, an approach has been developed for filling in missing values based on exponential smoothing with adaptive coefficients determined by Fibonacci numbers. This method effectively accounts for both short-term and long-term patterns in the data, contributing to a more accurate reconstruction of lost values. This article examines a method for recovering missing data based on a combined exponential smoothing model, applied to hourly energy consumption data for the period 2016–2018. The proposed approach utilizes a regression model in which the regressors are the values of the exponential moving average, determined using smoothing coefficients and window sizes based on Fibonacci numbers. This approach effectively accounts for both new and older information by adapting weight coefficients for more accurate recovery of missing values. It has been found that when small Fibonacci numbers are used to determine the size of the sliding window and weight coefficients, the formation of exponentially smoothed values is primarily influenced by the most "recent data" (the latest, most recently obtained values). The use of Fibonacci numbers to determine smoothing parameters in the exponential smoothing method allows for the adaptive consideration of both short-term and long-term trends in time series. The proposed data reconstruction model is based on combined forecasting using six exponential smoothing models, whose parameters correspond to specific Fibonacci numbers. The combined model is built through regression analysis, enabling the adaptive evaluation of weight coefficients at each forecasting step.*

Keywords: *mathematical modeling, data processing, data recovery, exponential smoothing, Fibonacci numbers.*

О.М. Терентьев, В.О. Дуда

Інститут телекомунікацій і глобального інформаційного простору НАН України,
м. Київ, Україна

МЕТОД ВІДНОВЛЕННЯ ПРОПУСКІВ У ДАНИХ НА ОСНОВІ КОМБІНОВАНОЇ МОДЕЛІ ЕКСПОНЕНЦІЙНОГО ЗГЛАДЖУВАННЯ

***Анотація.** Обробка та аналіз даних часто супроводжуються проблемою пропущених значень, що може значно впливати на точність моделей прогнозування та ухвалення рішень. Однією з основних причин виникнення пропусків у даних щодо енергоспоживання є періодичне відключення систем, що безпосередньо займаються збором та передачею інформації. Такі перебої можуть спричинити втрату частини даних, що ускладнює їх подальшу обробку та аналіз. Тому розроблено підхід до заповнення пропущених значень, заснований на експоненційному згладжуванні з адаптивними коефіцієнтами, що визначаються за числами Фібоначчі. Такий метод дозволяє ефективно враховувати як короткострокові, так і довгострокові закономірності в даних, що сприяє більш точному відновленню втрачених значень. В даній статті розглядається метод відновлення пропущених даних на основі комбінованої моделі експоненційного згладжування, застосованої до погодинних даних енергоспоживання за період 2016–2018 років. Запропонований підхід використовує регресійну модель, у якій регресорами виступають значення експоненційного ковзного середнього, визначені за допомогою коефіцієнтів згладжування та розмірів ковзного вікна, що базуються на числах Фібоначчі. Такий підхід дозволяє ефективно враховувати як нову, так і старішу інформацію, адаптуючи вагові коефіцієнти для точнішого відновлення пропущених значень. Використання чисел Фібоначчі для визначення параметрів згладжування у методі експоненційного згладжування дозволяє адаптивно враховувати як короткострокові, так і довгострокові тенденції в часових рядах. Запропонована модель відновлення даних ґрунтується на комбінованому прогнозуванні з використанням шести моделей експоненційного згладжування, параметри яких відповідають певним числам Фібоначчі.*

***Ключові слова:** математичне моделювання, обробка даних, відновлення даних, експоненційне згладжування, числа Фібоначчі.*

<https://doi.org/10.32347/2411-4049.2025.1.125-131>

Вступ

Обробка та аналіз даних часто супроводжуються проблемою пропущених значень, що може значно впливати на точність моделей прогнозування та ухвалення рішень. Однією з основних причин виникнення пропусків у даних щодо енергоспоживання є періодичне відключення систем, що безпосередньо займаються збором та передачею інформації. Такі перебої можуть спричинити втрату частини даних, що ускладнює їх подальшу обробку та аналіз.

Тому розроблено підхід до заповнення пропущених значень, заснований на експоненційному згладжуванні [1] з адаптивними коефіцієнтами, що визначаються за числами Фібоначчі [2]. Такий метод дозволяє ефективно враховувати як короткострокові, так і довгострокові закономірності в даних, що сприяє більш точному відновленню втрачених значень.

Запропонований метод може знайти застосування у широкому спектрі завдань, пов'язаних із аналізом часових рядів, зокрема в енергетичній галузі [3], де точність прогнозування та відновлення даних має вирішальне значення.

Аналіз останніх досліджень і публікацій. Схоже дослідження було проведено в роботі "Прогнозування короткочасних даних для крипторинку" [4], де розглядалося застосування методів експоненційного згладжування для прогнозування ціни криптовалют. У цій роботі оцінювалася точність прогнозування за допомогою показників RMSE, MAPE та аналізу тренду. Дослідження базувалося на даних криптобіржі Binance для тікеру BTC-USDT і передбачало тестування різних параметрів розміру вікна та коефіцієнта згладжування. Результати показали, що метод простого експоненційного згладжування дає найкращі результати у вибраному проміжку часу. Це підтверджує ефективність експоненційного згладжування для роботи з часовими рядами, що і стало основою для застосування аналогічного підходу у даній роботі для відновлення пропущених значень у даних енергоспоживання.

Коефіцієнти згладжування на основі чисел Фібоначчі. Використання чисел Фібоначчі [5] для визначення параметрів згладжування забезпечує адаптивність моделі, дозволяючи враховувати як короткострокові, так і довгострокові тенденції у даних. Такий підхід сприяє точнішому відновленню пропущених значень, оскільки він поєднує гнучкість експоненційного згладжування з математично обґрунтованим вибором коефіцієнтів.

На основі чисел Фібоначчі обчислюються коефіцієнти згладжування для методу експоненційного згладжування.

Загальна формула для обчислення згладженого значення EWM_t (Exponential Weighted Moving Average) має вигляд [1]:

$$EWM_t = (1 - \alpha) \cdot EWM_{t-1} + \alpha \cdot X_t,$$

де EWM_t – середнє значення за методом експоненційного згладжування у момент часу t ,

X_t – значення вхідного ряду даних у момент часу t ,

EWM_{t-1} – середнє значення у момент часу $t - 1$,

α – це коефіцієнт згладжування, $0 \leq \alpha \leq 1$, значення якого обчислюються за формулою [1]:

$$\alpha = \frac{2}{SPAN + 1},$$

де SPAN – це показник, який прирівнюється відповідному значенню числа Фібоначчі. SPAN в перекладі з англійської означає "інтервал", "ширина" або "діапазон".

За своїм сенсом коефіцієнт згладжування (рис. 1) вказує на те, який вклад робить попереднє значення у обчислення загального значення ковзного середнього.

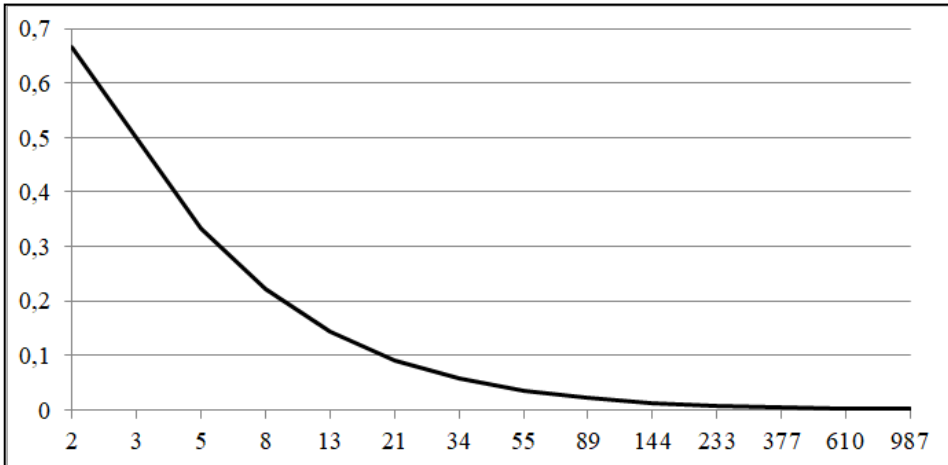


Рис. 1. Вісь OX – значення числа Фібоначчі (параметр SPAN), вісь OY – значення коефіцієнта згладжування α

За результатами вагових коефіцієнтів, варіювання значення SPAN впливає на обчислення значення коефіцієнта згладжування α , яке показує який внесок у формування EWM робить нове значення. Відповідні вагові коефіцієнти наступних лагових значень спадають експоненційно [1] за формулою $\alpha \cdot (1 - \alpha)^n$ і послідовно наближаються до нуля (табл. 1).

Таблиця 1. Значення коефіцієнтів значень лагу

n	Лаг	Коефіцієнт згладжування
0	t	$\alpha = 0,222(2)$
1	t-1	$\alpha \cdot (1 - \alpha) = 0,17284$
2	t-2	$\alpha \cdot (1 - \alpha) \cdot (1 - \alpha) = 0,134431$
3	t-3	$\alpha \cdot (1 - \alpha)^3 = 0,104557$
4	t-4	$\alpha \cdot (1 - \alpha)^4 = 0,081322$
5	t-5	$\alpha \cdot (1 - \alpha)^5 = 0,063251$
6	t-6	$\alpha \cdot (1 - \alpha)^6 = 0,049195$
7	t-7	$\alpha \cdot (1 - \alpha)^7 = 0,038263$
8	t-8	$\alpha \cdot (1 - \alpha)^8 = 0,02976$
9	t-9	$\alpha \cdot (1 - \alpha)^9 = 0,023147$
...
66	t-66	$\alpha \cdot (1 - \alpha)^{66} = 0,000000014$
67	t-67	$\alpha \cdot (1 - \alpha)^{67} = 0,000000011$
68	t-68	$\alpha \cdot (1 - \alpha)^{68} = 0,000000008$
69	t-69	$\alpha \cdot (1 - \alpha)^{69} = 0,000000007$
70	t-70	$\alpha \cdot (1 - \alpha)^{70} = 0,0000000050$

Метод відновлення даних на основі комбінованого прогнозу на основі результатів моделювання методом експоненційного згладжування. Для відновлення відсутніх даних використовується метод, що включає побудову шести моделей експоненційного згладжування [6] (M1–M6) з параметром SPAN, що відповідає числам Фібоначчі: 13, 144, 233, 377, 610 і 987.

Після чого будується модель комбінованого прогнозу, на основі прогнозних значень, отриманих на математичних моделях М1–М6, із використанням моделі регресії наступного вигляду

$$M_{COMBINE} = a_0 + \sum_{i=1}^6 a_i \cdot M_i .$$

Значення коефіцієнтів моделі, $a_0, a_1, a_2, a_3, a_4, a_5, a_6$, оцінюються адаптивно динамічно на кожному кроці, що вказано в таблиці 2.

Таблиця 2. Приклад побудови таблиці оцінок параметрів моделі, на 1485-й ітерації алгоритму

Змінні	Оцінка параметра моделі	Стандартна похибка	t-значення Стьюдента	p-значення Стьюдента
a_0	0,0109	0,00549	1,98	0,0472
a_1	0,19712	0,00526	37,44	< 0,0001
a_2	0,12885	0,00642	20,08	< 0,0001
a_3	0,06492	0,0076	8,54	< 0,0001
a_4	0,07883	0,01309	6,02	< 0,0001
a_5	-0,02211	0,02373	-0,93	0,3516
a_6	0,34387	0,02814	12,22	< 0,0001

В таблиці 3 наведено результати моделювання, щодо відновлення пропусків в даних, на основі даних погодинного енергоспоживання від компанії ДТЕК за період 2016-2018 років. В якості критерію оцінювання якості роботи методу використовувалася статистика MAPE (mean absolute percent error) – середня абсолютна похибка в процентах.

Таблиця 3. Значення похибки MAPE при заповненні пропусків в даних із використанням запропонованої моделі

Погодинний часовий ряд	Процент пропусків в даних, %			
	5	10	15	20
Дані ДТЕК за 2016 рік	5,03	6,53	7,57	8,92
Дані ДТЕК за 2017 рік	5,08	5,97	7,58	9,87
Дані ДТЕК за 2018 рік	5,32	7,28	9,54	11,61

Як можна побачити з отриманих результатів, наведених в таблиці вище, чим більше відсоток пропусків даних, тим гірша статистика MAPE, тобто збільшується розходження між реальним та відновленим значенням пропуску. Для випадків наявності 5% пропусків середнє значення похибки MAPE варіюється від 5,03 до 5,32% та поступово збільшується при зростанні кількості пропусків, та у випадку наявності 20% відсутніх даних сягає значення MAPE від 8,87 до 11,61%.

На рисунку 2 наведений частковий приклад реальних та змодельованих значень пропусків даних щодо задачі енергоспоживання.

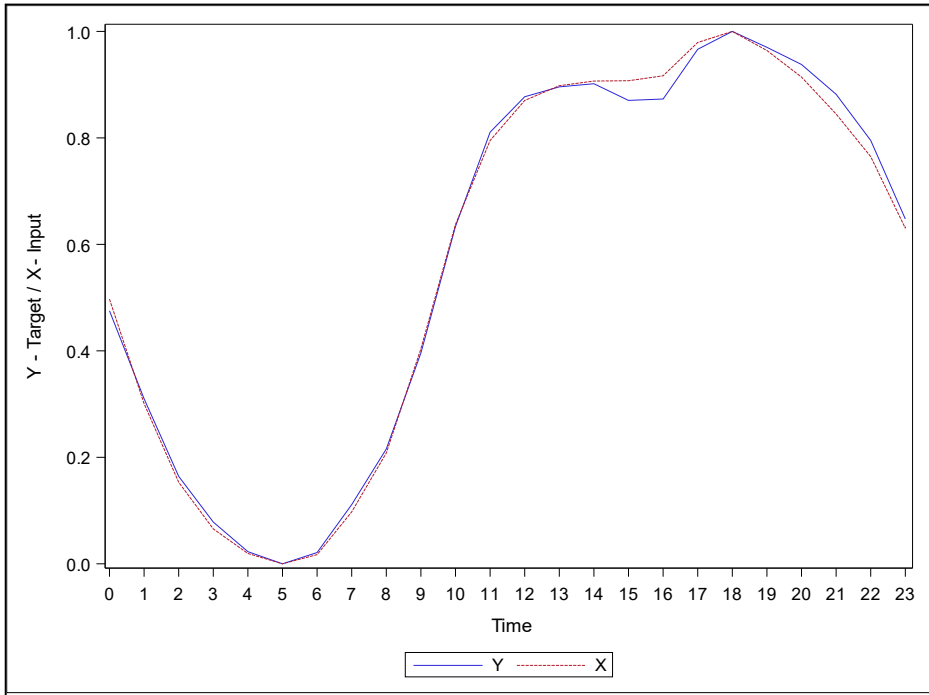


Рис. 2. Приклад відновлення пропусків даних у період з 14 до 18 години. Вісь ОХ – години, вісь ОУ – реальне та прогнозне значення погодинного енергоспоживання. Суцільна лінія – реальні дані, штрихова лінія – дані з відновленими пропусками

Висновки

З’ясовано, що у випадках, коли використовуються малі значення чисел Фібоначчі для формування розміру ковзного вікна та вагових коефіцієнтів, на формування експоненційно згладжених значень в першу чергу впливають найбільш “свіжі дані” (останні, нещодавно отримані).

В той час, як по мірі збільшення значень чисел Фібоначчі, що використовуються для формування розміру ковзного вікна та вагових коефіцієнтів, відбувається перетікання ступеня важливості та загального впливу від нещодавно отриманих даних до більш “застарілих”, але це дозволяє в свою чергу враховувати при обчисленнях більш довгу історію, щодо накопичених даних про процес, що досліджується.

Використання чисел Фібоначчі для визначення параметрів згладжування у методі експоненційного згладжування дозволяє адаптивно враховувати як короткострокові, так і довгострокові тенденції в часових рядах.

Запропонована модель відновлення даних ґрунтується на комбінованому прогнозуванні з використанням шести моделей експоненційного згладжування, параметри яких відповідають певним числам Фібоначчі. Комбінована модель будується шляхом регресійного аналізу, що дозволяє адаптивно оцінювати вагові коефіцієнти на кожному кроці прогнозування.

Експериментальні результати, отримані на даних погодинного енергоспоживання компанії ДТЕК за 2016–2018 роки, показують, що якість заповнення пропущених значень залежить від відсотка відсутніх даних. При

5% пропусків середня абсолютна похибка MAPE варіюється в межах 5,03–5,32%, тоді як при 20% відсутніх даних вона зростає до 8,87–11,61%. Це підтверджує, що метод ефективно працює для заповнення незначної кількості пропущених значень, але втрачає точність при великій кількості відсутніх даних.

REFERENCES

1. Brown, R.G. (1956). Exponential Smoothing for Predicting Demand. Cambridge, Massachusetts.
2. Tiner, J.H. (2004). Exploring the World of Mathematics: From Ancient Record Keeping to the Latest Advances in Computers. New Leaf Publishing Group.
3. Cao, Y., Yu, J., Zhong, R. et al. (2025). Forecasting Renewable energy and electricity consumption using evolutionary hyperheuristic algorithm. *Sci Rep*, 15, 2565. <https://doi.org/10.1038/s41598-025-87013-8>
4. Duda, V., & Rolik, O. (2023). Forecasting short-term data for the crypto market. *Adaptive Automatic Control Systems*, 1(42), 141–152 (In Ukrainian). [Дуда, В., & Ролік, О. (2023). Прогнозування короточасних даних для крипторинку. *Адаптивні Системи Автоматичного Управління*, 1(42), 141–152]. <https://doi.org/10.20535/1560-8956.42.2023.279107>
5. Grigas, A. (2013). The Fibonacci sequence. In Senior Thesis [Thesis]. Liberty University.
6. Risteski, D., Kulakov, A., & Davcev, D. (2005). Single exponential smoothing method and neural network in one method for time series prediction. *IEEE Conference on Cybernetics and Intelligent Systems*, 2, 741–745. <https://doi.org/10.1109/iccis.2004.1460680>

Стаття надійшла до редакції 14.10.2024 і прийнята до друку після рецензування 23.01.2025

The article was received 14.10.2024 and was accepted after revision 23.01.2025

Терентьєв Олександр Миколайович

доктор технічних наук, доцент, провідний науковий співробітник, Інститут телекомунікацій і глобального інформаційного простору НАН України

Адреса робоча: бульв. Чоколівський, 13, Київ, Україна, 03186

ORCID ID: <https://orcid.org/0000-0002-4288-1753> **e-mail:** o.terentiev@gmail.com

Дуда Володимир Олександрович

аспірант, Інститут телекомунікацій і глобального інформаційного простору НАН України

Адреса робоча: бульв. Чоколівський, 13, Київ, Україна, 03186

ORCID ID: <https://orcid.org/0009-0002-4278-4635> **e-mail:** dudavolodimir@gmail.com